

TUTTI I MODI PER VALUTARE LA RICERCA

di Daniele Checchi

Il problema del come valutare la ricerca scientifica appassiona sia i metodologi sia i comuni ricercatori, anche perché qualunque sia il metodo adottato esso verrà giudicato come inadeguato da parte di qualcuno. Un lavoro di ricerca può essere valutato per l'originalità, il rigore metodologico, la sistematizzazione della conoscenza pregressa, lo spostare in avanti la frontiera della ricerca, il travalicare i confini disciplinari, oltre che per l'impatto sulle pratiche sociali e la capacità di modificare il pensiero comune. Spesso i lavori di ricerca, siano essi libri, articoli su riviste scientifiche o contributi in raccolte collettanee, eccellono lungo una o più di queste dimensioni, ma sono rari i casi di lavori che risultino eccellenti secondo tutte le dimensioni. Una circostanza da tenere presente mentre si discute di come riformare il nostro modello di valutazione della ricerca universitaria.

L'esperienza ci insegna che risulta difficile (se non impossibile) costruire metriche valutative che tengano simultaneamente conto di tutte queste dimensioni, e si preferisce per questo ricorrere alla valutazione tra pari (*peer review*) ritenendo che un ricercatore competente della materia possa ponderare al meglio le diverse dimensioni. Ma anche gli umani hanno le loro preferenze, che spesso non sono coincidenti né con quelle degli autori sottoposti a valutazione né si

accordano tra loro. Chi abbia esperienza di invio di articoli a riviste che applichino la valutazione delle proposte di pubblicazione sa benissimo che sono rari i casi in cui due (o più) valutatori anonimi, selezionati secondo il principio della competenza disciplinare, esprimano pareri analoghi sui lavori stessi.

Lo stesso problema si pone infatti nel disegno degli esercizi di valutazione su scala nazionale, come per esempio quelli condotti da Hefce-Higher education funding council for England nel caso inglese con il Ref-Research excellence framework, o da Anvur nel caso italiano con la Vqr-Valutazione della qualità della ricerca. Il primo esercizio ha valutato 190.962 prodotti di ricerca realizzati nel quinquennio 2008-13 nelle università inglesi, il secondo ha preso in considerazione 114.431 prodotti di ricerca realizzati nel quadriennio 2011-14 nelle università e nei centri di ricerca pubblici italiani.

I due esercizi valutativi sono metodologicamente diversi: il Ref inglese lasciava all'autonomia delle università di scegliere quali ricercatori sottoporre a valutazione (anche se commisurava il finanziamento conseguente al numero di valutati), la Vqr italiana richiedeva invece che tutti i ricercatori si sottoponessero a valutazione. Il Ref inglese ha utilizzato come unico metodo valutativo la valutazione tra pari realizzata da un migliaio di valutatori, raccolti in 36 *panel*,

che hanno lavorato quasi a tempo pieno per circa un anno; la Vqr italiana ha combinato il metodo della valutazione tra pari con l'utilizzo di indicatori bibliometrici (numero di citazioni e visibilità delle riviste), impegnando 436 valutatori raccolti in 16 *panel*, che a loro volta hanno coinvolto 17 mila valutatori anonimi. L'avvalersi o meno di indicatori citazionali produce un differenziale di costo dell'esercizio non trascurabile: il costo del Ref inglese è stato stimato pari a 246 milioni di sterline, di cui 217 impiegati alla sola valutazione della ricerca (con un costo procapite di 5.500 euro per ricercatore valutato), il costo della Vqr è stato stimato in quasi 15 milioni di euro (con un costo procapite di 242 euro per valutato). In entrambi gli esercizi gli esiti della valutazione sono stati utilizzati per distribuire una parte del finanziamento pubblico agli atenei.

A differenza che nel caso italiano, il Ref inglese ha reso pubblici i lavori che sono stati sottoposti a valutazione (quattro per ogni ricercatore valutato), e questo ha permesso a un gruppo di ricercatori (cui ha contribuito anche chi scrive) di applicare esclusivamente gli indicatori citazionali per valutare questi stessi lavori. In questo modo è stato possibile realizzare un esercizio controfattuale, ovvero ci si è chiesti cosa sarebbe accaduto se i prodotti sottoposti al Ref fossero stati valutati dai *panel*

italiani utilizzati per la Vqr. I risultati sono sorprendenti: la correlazione tra la graduatoria dei dipartimenti prodotta dal Ref reale e quella prodotta dal Ref controfattuale è pari a 0,63 (essendo 1 il valore in caso di perfetto allineamento), e tale valore sale a 0,76 quando si considerino le discipline scientifiche (dove le tradizioni disciplinari attribuiscono maggior peso all'impatto citazionale).

Al di là quindi dell'aspetto relativo ai costi della valutazione, questa ricerca mette in luce come nelle aree scientifiche (quelle che nel gergo universitario vengono chiamate "aree bibliometriche") l'uniformità dei registri valutativi sia più pronunciata, al punto che un algoritmo può replicare con un discreto livello di approssimazione il lavoro di un gruppo di valutatori, per lo meno quando si tratti di ordinare dipartimenti che raccolgono gruppi di ricercatori.

Questa uniformità viene interpretata come conformismo da molti studiosi che operano nelle scienze sociali o umanistiche, senza però tener conto del fatto che la valutazione espressa dai pari contiene elementi di disaccordo talvolta così pronunciati da rendere per lo più ambigui gli ordinamenti dei dipartimenti che ne conseguono.

Ordinario di Economia all'Università
Statale di Milano e membro
del Consiglio direttivo di Anvur

NELLE MATERIE
SCIENTIFICHE
UN ALGORITMO
PUÒ DARE LE
GARANZIE DELLA
PEER REVIEW